ChinaXiv合作期刊

基于特征航迹简化模型的中心航迹提取研究 *

王莉莉，彭 勃

(中国民航大学 天津市空管运行规划与安全技术重点实验室, 天津 300300)

摘 要：针对现有中心航迹提取方法存在的问题：进近转弯点容易缺失，2。数据量庞大、运算耗时，有些航迹点没有代表性。根据“以直代曲”思想简化航迹点数据，然后基于特征航迹的航向进行聚类，求得平均航向并得到中心航迹定位点，最后使用 B 样条曲线拟合得到中心航迹。仿真结果表明，使用新方法航迹点数量减小 78.92%，同时准确地得到中心航迹，解决了转弯点缺失和耗时的问题，验证了新方法对于中心航迹的提取具有准确性和有效性。

关键词：中心航迹提取；特征航迹；进近转弯点；航迹聚类；以直代曲

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2017.07.0663

Research on central track extraction based on simplified track model

Wang Lili, Peng Bo

(*Tianjin Key Laboratory of Atc Operation Planning & Safety Technology*, *Civil Aviation Administration of China*, *Tianjin* 300300, *China*)

Abstract: In view of the existing problems of the central track extraction method : 1. approach turn points are easy to be missing; 2. the amount of data is huge and the computation time is time-consuming; 3. some track points are not representative. Firstly, simplify the track point data according to the idea of "replace curve by straight"in this paper, then cluster base on the characteristic track course, obtain the average heading and the center track locating point, finally, use the B spline curve to fit and obtain the central track, simulation results show that the number of track points decreases by 78.92% using the new method, at the same time, obtain the central track accurately, and solved the problem of turning point missing and time consuming, prove that the new method is accurate and effective for the extraction of the central track.

Key Words: central track extraction; feature track; approach turning point ; track clustering; replace curve by straight

chinaXiv:201805.00229v1

0 引言

通过分析航空器的历史雷达数据，航空公司能够客观地评价飞行员的飞行品质、飞机性能和航班的运行情况，进而科学地调整飞行计划[3~8]；对管理单位来说，可以利用航迹数据研究飞行程序的设计与优化以及噪声预测[9,10]。然而空中交通流量与日俱增，航空器的航迹数据也呈几何级数增长，航迹数据的数量十分庞大，人工处理这些数据显然是不实际，因此如何从海量航迹数据中挖掘出有价值的信息越来越受到民航领域研究者的重视，国内外很多学者对航迹聚类进行了相关研究，赵元棣等[3]将航迹视为以机场雷达为原点建立的立体坐标系下的离散点云集合，利用计算几何的骨架提取算法，求得代表点，经过处理得到中心航迹并识别出中心航迹的特征点；王洁宁等人[4]分析了国内外现有的航迹聚类算法，认为现有的航迹聚类算法忽略了时间信息，而航迹点的分布与进场时间有者密切的联系，提出了时间空间的聚类方法；王涛波等人[6]研究了基本的聚类算法，认为其在航迹聚类中存在容易陷入局部最优搜索的缺点，提出了采用遗传模拟退火算法对模糊 C 均值聚类算法进行优化；徐涛等人[8]提出了基于航迹间面积加权的航迹相似性度量的 CURE 算法，使聚类效果更精确；Rehm 等人[11]直接利用已有航迹定义其相似性，针对同一机场不同跑道的进近程序聚类。Gariel 等人[12]进行了雷达航迹聚类，并对航迹运行进行实时监控。Kenneth 等人[13]针对机场的繁忙区域，采用基于网格的聚类方法对不同机场的航迹进行分析，得到其不同机场的进离场中心航迹。

目前，在对航迹聚类的研究中，进近转弯点中心航迹的提取和拟合是研究的难点。在对进近航迹进行聚类，提取中心航迹时，最终结果往往是直线航段能够得到中心航迹，而转弯航段则容易缺失。因此，如何对进近转弯点的历史航迹进行有效的中心航迹提取与拟合具有很好的理论价值和实际意义，本文提出基于特征航迹简化模型的中心航迹提取方法，根据“以直代曲”思想简化航迹点得到特征点，然后基于特征航迹的航向

基金项目：国家自然科学基金委员会与中国民用航空局联合资助项目（U1633124）；民航局科技创新引导项目（20150231）

作者简介：王莉莉（1973-），女，教授，主要研究方向为空中交通管理、空域规划（）；彭勃（1991-），男（满族），硕士研究生，主要研究方向为航迹聚类分析.

ChinaXiv合作期刊
chinaXiv:201805.00229v1

进行聚类，对聚类结果中的每一类别的特征航迹求得平均航向同时得到中心航迹定位点，最后使用 B 样条曲线拟合得到平滑的中心航迹，本文通过对进近转弯航段真实的历史航迹数据进行实验分析，验证了该方法的准确性和有效性。

1 雷达数据处理和进近过程描述

本文要处理的数据源是原始二次雷达数据，首先对原始二次雷达数据进行解码转换成全文本明码类型，然后再将所有含有有用信息的文本条目按关键字逐一检出，并将所需字段转换成程序需要的二进制整数或浮点数格式，数据单位保留不变，最后整理得到的规范航迹数据形式如表 1 所示。

表 1 整理得到的二次雷达数据

时间	纬度	经度	高度	速度	航向
11:40:26	112.2890	32.1376	274	279	169.00
11:40:30	112.2890	32.1376	274	226	183.00
11:40:34	112.2870	32.1293	426	247	175.00

虽然可以从该表的雷达数据中获取航向信息，但是雷达数据每 4 秒钟采集一次，对于 D 类航空器，转弯速度较大，4 s 时间过长，不能准确地刻画出航空器的航向变化规律，所以在本文第 2 节中重新定义了特征点、特征特征航迹以及航向变化等几个概念。由于二次雷达返回的是经纬度信息，只有转换成大地坐标才能准确的反映出飞机的实际运行轨迹，因此以机场的导航台为坐标原点，建立直角坐标系，将返回航迹的经纬度数据转换为大地坐标，坐标转换得到直角坐标系下的横纵坐标（单位为千米）如表 2 所示。

表 2 经纬度坐标转换成平面坐标

	横坐标	纵坐标
1	00173.93750	00244.25000
2	00173.56250	00243.62500
4	00173.25000	00243.00000

将原始数据经过坐标转换处理后用 MATLAB 作图得到该进近程序在实际运行中的情况，如图 1 所示，筛选截取其中三段转弯部分分别命名为航段 1、航段 2 和航段 3。该程序的飞行过程大致如下：脱离航路后向台飞行，在导航台上方转弯，此为航段 1 的部分；然后直飞下一个导航台（IAF），过台转弯到下一个导航台，然后再过台转弯，此为航段 2 的部分；此后进入基线转弯程序直至落地，此为航段三的部分。由得到的数据显示，平均一条航迹一个转弯过程大约有 20 个点迹，该机场一年的起降架次大约三十多万，也就是说一年的进近转弯数据累计可达进四百万。这样庞大的数据量用计算机处理起来不仅耗时，而且需要占用非常大的内存，不经济。因此需要一种方法既可以在很大程度上减少航迹数据，又可以保证航迹的准确性。

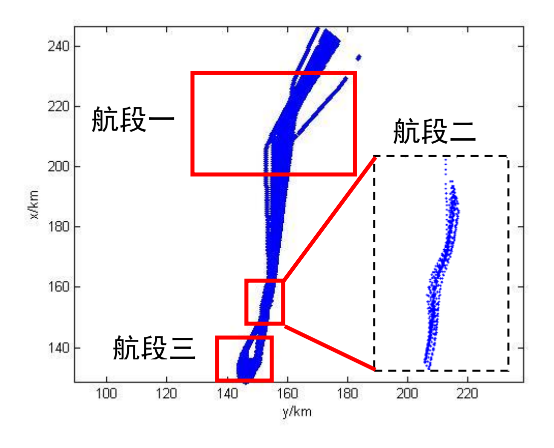


图 1 进近航段

2 基于特征航迹的简化模型

在数学和物理学科中有“以直代曲”的思想，认为一条曲线可以用在曲线上取点连成的折线段近似表示。对于转弯航迹来说，也可以利用这样的思想来做处理：当连续的某几段航迹走向大致相同时，可以用这几段连续航迹的起点和终点间的直线段来近似表示。本文把这里的起点和终点称为特征点，特征点间的直线段称为特征航迹。为了保证准确性，使特征航迹尽可能地接近原航迹，本文希望特征点是表示航迹走向有较大变化的点。假设有如图 2 中由 P_1 到 P_8 的航迹点构成的一段转弯航迹，可以看出 P_3 和 P_6 航迹点两侧航迹走向明显不同，且差异较大，所以这两点应该是特征点。由 CP_1 到 CP_4 组成的特征航迹可以体现出原航迹的趋势。

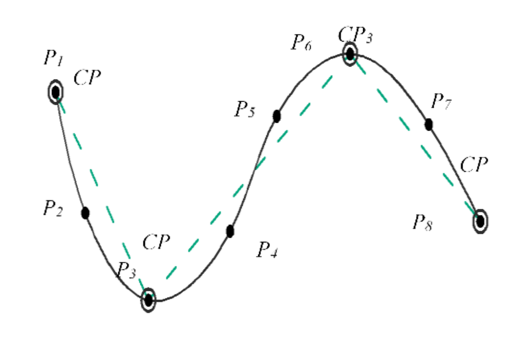


图 2 以直代曲原理图

在所有航迹点中选择特征点，而特征点的最佳选择要遵循两个原则：准确性原则和最简原则。准确性原则是使特征航迹与原航迹间的差异尽可能小，而最简原则是希望特征点的个数尽量少。这两个原则在选取特征点的过程中是相互对立的，只有综合考虑这两个原则的约束才能使得特征点的数量和计算精度达到要求，为了兼顾准确性原则和最简原则，使特征点的选取达到最优，本文引入特征航迹与原航迹的差异度量值——特征航迹差异矩阵 E 。E 的定义为原航迹的近似长度与特征航迹表征长度的差。其中原航迹的近似长度为两个航迹点间的直线长度，特征航迹表征长度包含三个部分，分别为：a）特征航迹长度 $d^{'}$；b）垂直差异量 d_p，即特征点之间的各原航迹点到特

ChinaXiv合作期刊
chinaXiv:201805.00229v1

征航迹的距离之和；c）角度差异量 d_a，如式（1）所示。

$$da=\begin{cases} l\times\sin\theta, 0\le\theta<90^o \\ l, 90^o\le\theta<180^o \end{cases} \tag{1}$$

其中：l 是各原航迹段长度，θ 是各原航迹段向量与特征航迹向量间的夹角，由式（1）可以看出，当 $\theta\le 90^o$ 时 d_a 等于该航段两端点垂直差异量之差，如图 3 所示，原航迹的长度为

$$d=|p_1+p_2|+|p_2+p_3|+|p_3+p_4|$$

特征航迹表征长度为

$d^{'}=|p_1+p_4|+(d_1+d_2)+|d_1+(d_2-d_1)+d_2|$ 则特征航迹差异量为 $E=d^{'}-d$。通过控制 E 的大小，使得简化选取的特征点结果符合两个原则。综上，新方法在解决处理大数据耗时问题的优势就体现在简化模型上，该简化模型可以在初始数据处理时将数据精简，从而减少算法耗时。

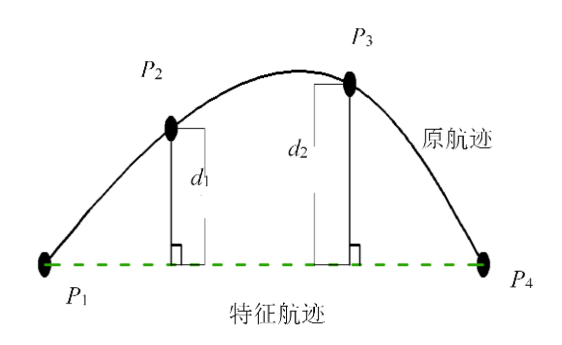


图 3　特征航迹差异量

3　基于航向的聚类

上文提到传统聚类算法在进近转弯点处的提取效果不好，原因如下：a)按照基本的聚类方法是对轨迹分布在数学上的评价指标（一般考虑距离、密度）进行聚类，没有对于航空器的运行过程进行考虑，最终结果往往只能缺失转弯部分而大致上体现出轨迹的线性走向;b)存在一些不能表征航空器实际运行规律的航迹点，要舍去此类航迹点，因此本文提出基于航向的聚类算法，通过该聚类算法可以将转弯过程体现出来，航空器的转弯轨迹在空间上表现的特点是转弯半径大，通常为 2~5 km，所以可以近似认为是飞几个直线段完成的转弯。同时转弯的本质是航向的变化，而航向的变化不是一蹴而就的，而是逐渐变化的过程。考虑以上情况，本文提出按照转弯航向对特征航迹进行聚类，将航向差异较小的航段认为是同一转弯阶段，通过求每一阶段的中心航迹从而最后还原出整个转弯过程的中心航迹。转弯阶段聚类的具体步骤是：将所有特征航迹的线段转换为向量形式，随机选择一条线段，计算它与其他向量间的夹角，通过角度阈值 θ，得到与其方向相似的向量作为同一簇，从而完成聚类。

4　中心航迹定位点的获取

传统的聚类算法将所有的航迹数据进行聚类，没有对航迹点进行分类进而对每一类中的航迹点聚类得到中心定位点，所以往往导致最后的中心定位点出现缺失或偏离轨迹的现象，为解决传统聚类算法得到的中心定位点不准确的问题，聚类后每一簇都代表了转弯的某一阶段，对每一簇的航迹求平均值，再将所有簇的数据结合起来可以反映该航段的历史航迹的分布趋势，也就是中心航迹。线段的平均值可以用某一条直线截取线段（下称截取线），计算截点的平均值。然而截取线不能随意选取，如果直线与簇内航迹间的夹角偏大，那么航迹左右两边则各有一段无法准确计算平均值，所以截取线应该与大部分航迹方向垂直。因此，引入平均向量表示簇内大部分航迹的方向。若选取的截取线与平均向量相垂直，则能够准确求得平均航迹的长度可以得到增加，中心航迹也更精确。在处理的时候需要将航迹用向量表示，平均向量的定义由式（2）给出。

$$\bar{v}=\frac{v_1+v_2+\ldots+v_{n-1}+v_n}{n} \tag{2}$$

得到平均向量后，求中心航迹只需要用垂直于平均向量的截取线扫过每一簇中的航段，然后求其交点的平均值即可，因为航迹是由航迹点间的线段构成。在求平均航迹时，应该考虑每一簇中的所有航迹点，用过每一个点的直线与航迹相交的点求平均值，得到点称为中心航迹定位点。连接所有中心航迹定位点便可以得到中心航迹。但并不是所有航迹点得到的平均值都能成为中心航迹定位点。

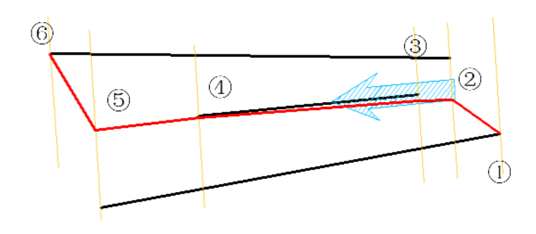


图 4　平均航迹原理图

如图 4 所示，假设某一簇中航迹如图分布，其中蓝色的箭头代表平均航向，橘黄色线段与平均航向垂直，红色轨迹表示通过计算每一航迹点的平均值得到的平均航迹。可以看出，如果对每一个点求平均值会使得结果存在很大的偏差。当截取线与航迹交点少于一定数量时，如图中的①和⑥，得到的中心航迹定位点的结果不可信，应该舍去这两个点。因此需要设置交点阈值，只有当交点个数大于某个值时，才认为此时得到的中心航迹定位点是可取的。其次为了使中心航迹更平滑，对于中心航迹定位点的间隔也需要进行控制。如图中的和③，因为②和③相隔太近所以跳过不取，因此图中的中心航迹应该为由定位点②、④和⑤构成。

5　样条曲线拟合

由上节程序得到了中心航迹定位点，将中心航迹定位点连线得到中心航迹。为了使最后得到的中心航迹平滑且符合动力学约束，本文选择使用三次 B 样条曲线进行连接。样条曲线在工程设计中应用广泛，是经过一系列给定点的光滑曲线，其在

chinaXiv:201805.00229v1

每个点处的一阶和二阶导数连续，具有连续、曲率变化均匀的特点，因此符合飞行性能。三次 B 样条曲线原理如下：

假设在区间（a，b）上给定一个分割：

$$\Delta: a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b \tag{3}$$

在（a，b）上的一个函数 $S(x)$ 称为差值三次样条函数，如果满足下列条件：

在每一个区间 $(x_{i-1}, x_i), i = 1,2,..,n$ 内 $S(x)$ 分别是三次多项式函数；在节点 $x_i (i = 1,2,...,n-1)$ 处成立：

$$s^{(k)}(x_i - o) = s^{(k)}(x_i + o), k = 0,1,2,...,n-1 \tag{4}$$

即小区间上的三次多项式函数，在拼接点处 x_i 具有二阶连续拼接。

满足差值条件 $y_i = S(x_i), i = 0,1,...,n$，当 $n = 3$ 时，则有三次 B 样条曲线的基函数如下：

$$\begin{cases} G_{0,3}(t) = \frac{1}{6}(-t^3 + 3t^2 - 3t + 1), \\ G_{1,3}(t) = \frac{1}{6}(3t^3 - 6t^2 + 4), \\ G_{2,3}(t) = \frac{1}{6}(-3t^3 + 3t^2 + 3t + 1), \\ G_{3,3}(t) = \frac{1}{6}t^3, \end{cases} \quad t \in [0,1] \tag{5}$$

三次 B 样条曲线段：

$$P_{0,3}(t) = \frac{1}{6}\begin{bmatrix} 1 & t & t^2 & t^3 \end{bmatrix} \begin{bmatrix} 1 & 4 & 1 & 0 \\ -3 & 0 & 3 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{bmatrix} \tag{6}$$

6 仿真分析

本文采用 MATLAB 8.3 环境编程进行仿真，选取某机场一套进近程序的转弯段航段三一周的航迹数据，数据样本中共有 15341 个航迹点，算法中需要设置的三个参数分别为：a） k，它决定了特征点的数量和特征航迹近似程度；b） $sita$，其决定了类的数量和聚类效果；c） $dist$，它控制中心航迹定位点间的距离和交点个数，决定了拟合航迹的光滑程度。首先选取不同的 k 值，对航迹点进行剔除，使得航迹点的选取符合准确性原则和最简原则，得到 k 值、剔除后剩余航迹点和剔除率的不同结果如表 3 所示。为验证剔除效果，从航段三中选取一条航迹数据，选取 k =0.20、0.21、0.22 时的剔除效果分别如图 5~7 所示，其中蓝色空心圆代表原始航迹点，红色实心点代表剔除后剩下的航迹点，从图中可以看出，选取 0.20 和 0.22 时的航迹点过于集中，存在两个点相邻的情况，而选取 0.21 时航迹点比较发散，分布较均匀，符合上文所述的两个原则，所以选取 k =0.21 进行航迹点剔除，对剔除后得到的航迹点进行聚类分析，选取不同角度阈值 $sita$ 进行聚类，如图 8 所示当选取 $sita = 20^o$ 时，将航段聚成八类，聚类效果很好，航段无缺失，然后选取不同的间距阈值 $dist$，选取 $dist$ =0.5、1.0 和 1.5 的效果图如图 9~11 所示，当选取 $dist$ =1.0 时航迹点间距较合适，航迹点比较分散，其他参数不变，取选取 k =0.20、0.21、0.22 时，算法耗时都在 5s 左右；为了验证本方法的优势，使用 KMEANS 算法与本方法对比，如图 12 所示：同样选取相同的数据进行聚类分析，算法耗时 44s，提取得到 17 个中心航迹定位点，用黑色十字表示，从图中可以看出 KMEANS 算法在直线段的提取效果很好，但是在转弯点处出现了缺失现象，尤其是第 10、11 个点，严重偏离中心航迹，得不到理想的中心航迹点提取效果，从而不能反映出航空器实际的运行规律，说明传统聚类算法在提取进近转弯点、解决算法耗时以及获取中心航迹的问题上存在不足。

随后对本方法得到的 17 个中心航迹点进行 B 样条拟合，选取插值间隔为 0.5 时得到 33 个控制点，表 4 中列出了部分控制点的坐标，随后将控制点依次连接得到最后中心航迹如图 13 所示，图中蓝色空心圆代表 17 个中心航迹点，红色十字代表使用 B 样条曲线得到的 33 个控制点。

表 3　不同阈值 k 得到的结果

阈值 k	0.20	0.21	0.22	...	0.25
剔除后航迹点	3327	3234	3167	...	2948
剔除率	78.31%	78.92%	79.36%	...	80.78%

表 4　B 样条曲线拟合得到的部分控制点

坐标点序列	1	2	3	...	33
控制点坐标 X	150.4979	150.3399	150.1492	...	147.6681
控制点坐标 Y	139.1093	138.1118	137.0123	...	141.9310

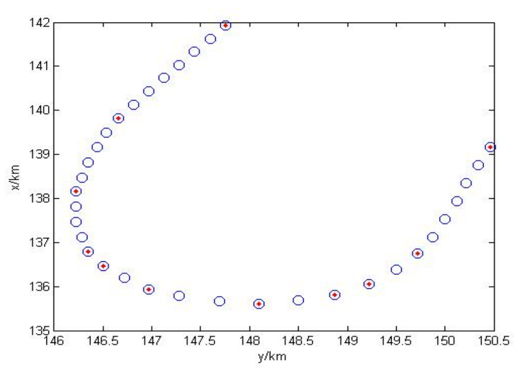


图 5　剔除率=0.20 的效果图

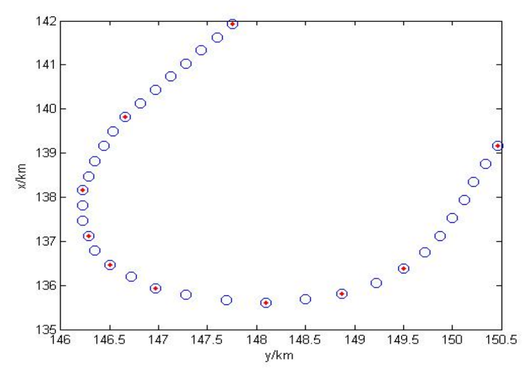


图 6　剔除率=0.21 的效果图

ChinaXiv合作期刊

chinaXiv:201805.00229v1

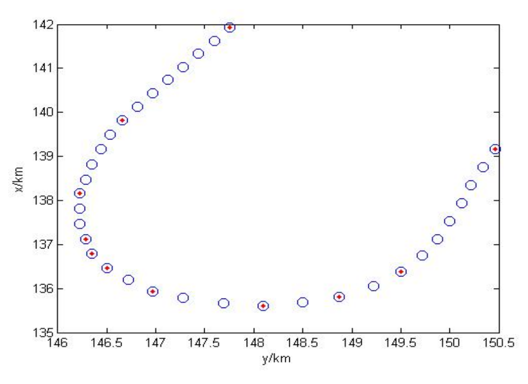


图 7　剔除率=0.22 的效果图

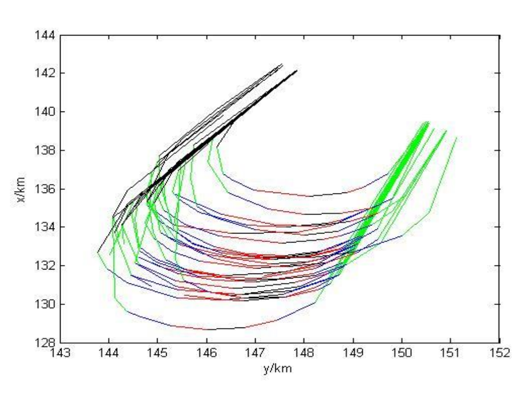


图 8　角度取 20 时聚成 8 类

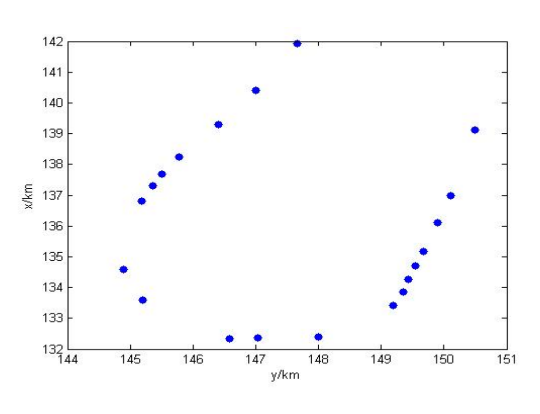


图 9　间距为 0.5 时得到 20 个航迹中心点

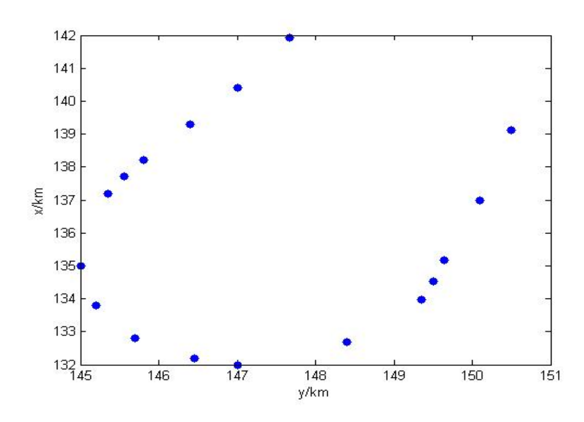


图 10　间距为 1.0 时得到 17 个航迹中心点

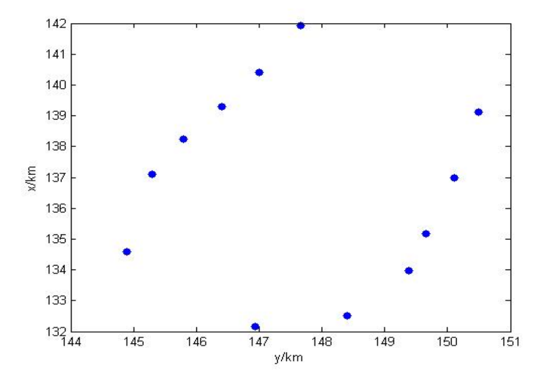


图 11　间距为 1.5 时得到 12 个航迹中心点

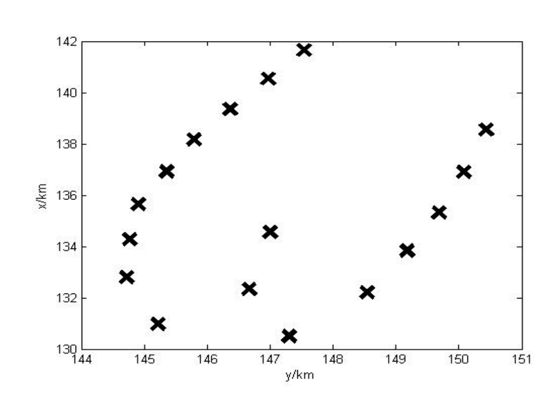


图 12　KMEANS 算法得到的 17 个航迹中心点

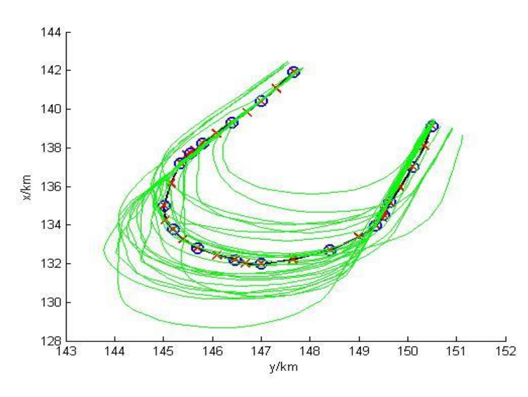


图 13　最后拟合得到的中心航迹

7　结束语

本文根据“以直代曲”思想提出基于特征航迹简化模型的中心航迹提取方法，分别经历了原始二次雷达数据转码、原始航迹点经纬度坐标转换、非特征航迹点的剔除和航迹点简化得到特征航迹点，进而对特征航迹聚类分析，提取得到中心航迹点定位点，使用三次 B 样条曲线拟合得到平滑的符合飞机飞行性能的中心航迹，通过对真实的二次雷达进近数据仿真同时与其他算法对比分析，结果表明本方法在转弯点处得到的中心航迹点无缺失现象，在处理相同数据时新方法耗时更少，解决了现有方法存在的问题，验证了新方法的准确性和有效性，对航迹聚类分析尤其是在转弯航段的中心航迹提取方面具有重要的意义与价值，下一步将围绕终端区内多条进场航班的航迹数据

ChinaXiv合作期刊

的中心航迹提取问题进行研究。

参考文献：

[1] Gudmundsson J, Thom A, Vahrenhold J. Of motifs and goals: mining trajectory data [C]// Proc of the 20th International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2012: 129-138.

[2] 吴笛，杜云艳，易嘉伟，等. 基于密度的轨迹时空聚类分析 [J]. 地球信息科学学报, 2015, 17 (10): 1162-1171.

[3] 赵元棣，孙禾，王洁宁，等. 终端区航迹簇的中心航迹提取方法研究 [J]. 图学学报, 2014, 35 (3): 379-386.

[4] 王洁宁，孙禾，赵元棣. 基于时间-空间的进场航迹聚类分析 [J]. 科学技术与工程, 2013, 13 (33): 10078-10081, 10088.

[5] 王超，徐肖豪，王飞. 基于航迹聚类的终端区进场程序管制适用性分析 [J]. 南京航空航天大学学报, 2013, 45 (1): 130-139.

[6] 王涛波，黄宝军. 改进模糊聚类算法在航迹分析中的应用 [J]. 中国安全科学学报, 2014, 24 (2): 23-27.

[7] 王超，王明明，王飞. 基于改进的模糊 C-means 航迹聚类方法研究 [J]. 中国民航大学学报, 2013, 31 (3): 14-18.

[8] 徐涛，李永祥，吕宗平. 基于航迹点法向距离的航迹聚类研究 [J]. 系统工程与电子技术, 2015, 39 (9): 2198-2204.

[9] 李永祥. 面向机场噪声预测的航迹聚类研究及应用 [D]. 天津：中国民航大学, 2015.

[10] 徐涛，陈雪蕊，等. 基于航迹聚类的终端区飞行程序轨迹表示 [J]. 四川大学学报：工程科学版, 2016, (06): 188-196.

[11] Rehm F. Clustering of flight tracks [C]// AIAA infotech @ Aerospace, Atlanta: AIAA 2010-3412.

[12] Gariel M, Srivastava A N, Feron E. Trajectory clustering and an application to airspace monitoring [J]. IEEE. Trans on Intelligent Transportation Syatems, 2011, 12 (4): 1 511-1 524.

[13] Kenneth Leiden, Stephen Atkins. Trajectory clustering for metroplex operations [C]// Proc of the 11th AIAA Aviation Technology, Integration, and Operation Conference. 2011.

[14] Lee J G, Han J. , et al. Trajectory clustering: a partition-and-group framework [C]// Proc of ACM SIGMOD International Conference on Management of Data. 2007: 593-604

chinaXiv:201805.00229v1